DOCUMENT RESUME

ED 223 688 TM 820 833

AUTHOR Gabriel, Roy M.
TITLE Post-MANOVA Data Analysis: Six Alternatives and Criteria for Selection.
PUB DATE Sep 79
NOTE 73p.; Paper presented at the Annual Meeting of the American Psychological Association (87th, New York, NY September 1-5, 1979).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Analysis of Variance; Criteria; Data; Data Analysis; Evaluation Methods; *Hypothesis Testing; Mathematical Models; *Multivariate Analysis; *Research Methodology; Selection; *Statistical Analysis
IDENTIFIERS *Post Hoc Methods

ABSTRACT

Although there is a healthy trend toward employment of multivariate statistical methods in the analysis of data from experiments in psychology and education which involve multiple dependent variables, there remains a need to demonstrate the wide variety of data analytic techniques available for post hoc analysis of statistically significant multivariate results. This paper attempts to: (1) clearly illustrate the use of six pertinent statistical analysis methods on a common sample data set; (2) compare the results both in terms of sensitivity to the treatment differences and interpretive clarity with respect to the research hypothesis; and (3) present guidelines to the applied researcher for selecting among the alternatives in a particular research context. Post hoc data analytic methods presented are: univariate F tests, univariate multiple comparison methods, step-down F tests, linear discriminant analysis, multiple profile analysis, and multivariate simultaneous test procedures. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED223688

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. M. Gabriel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Post-MANOVA Data Analysis

Six Alternatives and Criteria for Selection

Roy M. Gabriel

Division of Evaluation, Research and Assessment

Northwest Regional Educational Laboratory

Portland, Oregon

Presented at the 87th annual meeting of the

American Psychological Association;

New York, New York, September 1-5, 1979.

Post-MANOVA Data Analysis

Six Alternatives and Criteria for Selection

Abstract

Although there is a healthy trend toward employment of multivariate statistical methods in the analysis of data from experiments in Psychology and Education which involve multiple dependent variables, there remains a need to demonstrate the wide variety of data analytic techniques available for post hoc analysis of statistically significant multivariate results. This paper attempts to a) clearly illustrate the use of six pertinent statistical analysis methods on a common sample data set, b) compare the results both in terms of sensitivity to the treatment differences and interpretive clarity with respect to the research hypothesis, and c) present guidelines to the applied researcher for selecting among the alternatives in a particular research context. Post hoc data analytic methods presented are a) univariate F tests, b) univariate multiple comparison methods, c) step-down F tests, d) linear discriminant analysis, e) multiple profile analysis, and f) multivariate simultaneous test procedures.

Multivariate analysis of variance (MANOVA) methods have, through the advance of high speed computing facilities, become more widely represented in the applied literature in recent years. The appropriateness of their application is well known, as few empirical studies in education or psychology rely on a single criterion as evidence of treatment effects (Tatsuoka, 1973). The complexities of the techniques in their mathematical derivation or their multidimensional contrasts and assumptions notwithstanding, MANOVA represents a conceptually simple extension of well known univariate ANOVA techniques. No one has witnessed this point more saliently than Harris (1975) in his introductory remarks to his fine textbook:

> Each of the commonly employed techniques in multivariate statistics is a very straightforward generalization of some univariate statistical tool, with the single variable being replaced by a linear combination of several original variables. (xi)

Such works as these have brought to the applied researcher the interpretive simplicity of MANOVA. Other efforts have emphasized the experimentwise Type I error protection (e.g., Cramer & Bock, 1966) and the wealth of additional information (Gabriel & Glavin, 1978) afforded by MANOVA when multiple dependent variables are employed. When MANOVA results indicate statistically significant differences among levels of the factors under study, these differences are based upon the entire set of dependent variables. Frequently, such a result is too general to fully satisfy the interpretive interest of the researcher.

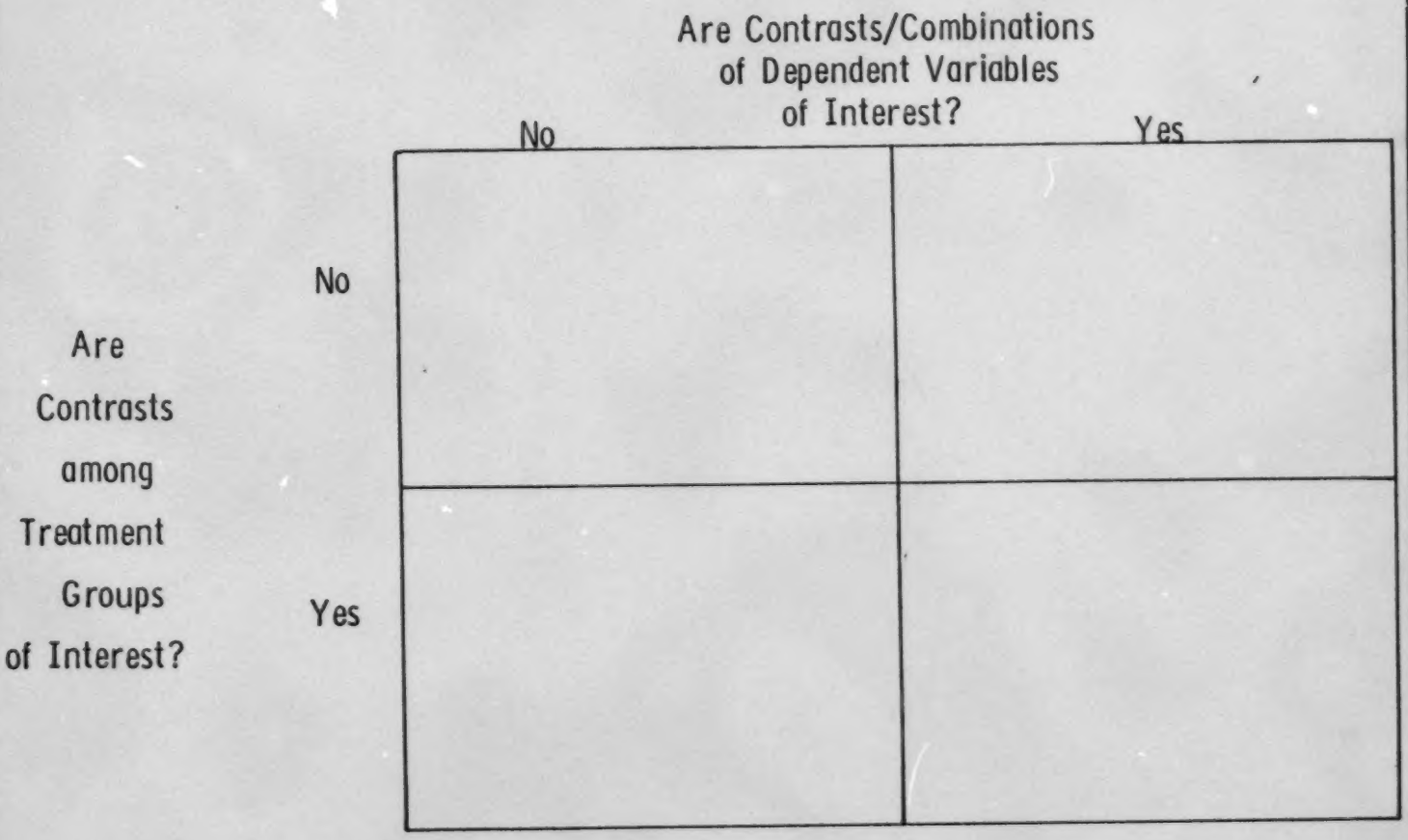
Although it provides a powerful, and Type I error-protected, beginning to the data analytic task, much more may be said of the effects of the treatment or differences among factor levels under study.

The purpose of this paper is to present a number of alternatives and criteria for selection of data analytic techniques which go beyond the omnibus MANOVA result to further probe the nature of the group differences present in the data. To refer to these as post hoc comparison techniques is semantically quite correct, but may fail to communicate the important considerations present in the multivariate context which are not present in the univariate context.

The first of these is that post hoc efforts in a multivariate analysis of variance may be conceived of as pertaining to two possible dimensions of the experimental design--(a) differences among treatment groups and (b) differential treatment effects on various dependent variables. Perhaps the first decision confronting the researcher is to determine whether only one of these, both dimensions separately, and/or combinations of both dimensions are of post hoc interest. For example, particular contrasts among the treatment levels may not be of interest, but individual criterion measures, rather than the entire set, may provide the interpretive focus. This, then, considers only the second "dimension" mentioned above. Alternatively, subsets of the package of dependent variables may be conceptually related and of interest as a set in particular contrasts among factor levels. This combines the factor level contrast dimension with the differential salience of criterion measures dimension. Just as these evoke

different research interests on the part of the researcher, they often point to different data analytic techniques. Though some techniques in wide use for these purposes may be applied to all dimensions of post hoc multivariate analyses, they are differentially powerful and often provide non-corroborative information.

Post-MANOVA Data Analysis



Related Literature

Most textbooks, while providing a thorough treatment of the overall MANOVA significance test procedures, do not present many alternatives for more specific analyses. Finn (1974) points out that univariate F tests indicate which dependent variables are most affected separately by the treatment manipulation, while being appropriately careful to say that their results should in no way be considered as partial tests of the multivariate hypothesis. Indeed they are not. Morrison (1976) presents the fully multivariate simultaneous test procedures (STP) analagous to the univariate Scheffe technique which, in the multivariate case, may be used to test all possible contrasts among dependent variables as well as groups. Bock (1975) presents prehaps the most comprehensive coverage of the utility of the step down F procedure in MANOVA. Harris (1975) suggests the interpretation and use of the leading discriminant function as a composite variable and suggests using it in contrasting treatment groups, using the sampling distribution of the greatest characteristic root (Heck, 1960) in conducting significance tests. Several other authors include detailed treatment of discriminant analysis as a multivariate technique of interest in its own right, not necessarily related to MANOVA (Tatsuoka, 1971b; Cooley and Lohnes, 1971).

That multivariate textbooks typically do not present a very thorough description of data analytic alternatives following the multivariate test is represented in Table 1, which lists a number of

TABLE 1

Coverage of Post MANOVA

Data Analytic Strategies in

Popular Multivariate Texts

	Univ. F	Step Down	Discrim. Analysis	Profile Analysis	Simult. Test	Univ. MC
Bock (1975)	*	*	*		*	
Cooley & Lohnes (1962)			*			
Cooley & Lohnes (1971)			*			
Finn (1974)	*	*	*			
Green (1978)	*		*			
Harris (1975)	*		*	*	*	*
Morrison (1967)	*			*	*	
Morrison (1976)	*		*	*	*	
Overall & Klett (1972)	*		*			
Press (1972)	*		*		*	
Tatsuoka (1971)			*			
Timm (1975)	*		*	*	*	*

the more popularly used references and indicates which of the six methods are covered in the text. As can be seen from the figure, only two are widely represented (discriminant analysis and univariate F tests). Even this portrays the situation more favorably than it exists, however, for rarely are the methods discussed as they relate to MANOVA and the multivariate hypothesis testing process. Instead they are presented as isolated techniques, often with little guidance as to when they might be most effectively employed in a research effort.

The methodological journal literature has been somewhat kinder to the applied researcher. In some earlier efforts at communicating the utility of multivariate methods, Cramer and Bock (1966) and Bock and Haggard (1968) discussed the usage of univariate F tests and step down F tests following MANOVA, pointing out the residualization of dependent variables in the latter and its advantages in eliminating the redundancy contained in sequential univariate tests. Still, with proper attention given to Type I error rate, they considered univariate F tests potentially useful in a post hoc effort. Hummel and Sligo (1971) presented Monte Carlo results comparing actual Type I error rates associated with the use of univariate F tests and multivariate STP following a significant overall mutivariate test. Their results illustrated the conservative nature of STPs and led them to the recommendation that they be used only when no other alternatives were available.

In a pair of articles, Stevens (1972, 1973) presented some derivational background and illustrated the use of a wider variety of available techniques. Discriminant analysis, step down F tests and STP were compared on sample data. Wilkinson (1975) contrasted four different methods of determining the relative importance of the dependent variables in accounting for treament group differences in MANOVA. Drawing heavily on the work of Darlington (1968) on the analagous issue in multiple linear regression, the author demonstrated on a sample data set that the various approaches yield different conclusions as to judgements of importance, and that no one of them by itself is sufficient in accounting for the relationship between treatment and response variables. In light of the complexity of multivariate hypotheses, neither of these conclusions is surprising.

Bird (1975) presented useful extensions of the univariate Tukey, Scheffe and Bonferroni-t multiple comparison techniques for partially or fully planned tests of contrasts in a multivariate experiment. He illustrates the benefits of a priori specification of contrasts by comparing results of these techniques with those of the STP associated with MANOVA. Berger (1978) further illustrated that STP are probably the method of choice only when the treatment group contrasts are made on the basis of combinations of dependent variables rather than individual dependent variables. Observing that the effects of the treatment on individual dependent variables are so frequently of interest to the multivariate researcher, he adds that these should be

specified in advance and a Bonferroni split of the familywise Type I error rate be utilized in conjunction with univariate F tests or multiple comparison techniques.

Spector (1977) illustrated the different results from a discriminant analysis and univariate ANOVAs on a multivariate data set, recommending that the latter are of interest in hypothesis testing research efforts while the former should be employed primarily in prediction and classification problems. Kaplan and Litrownik (1977) spoke primarily to an audience of applied behavioral researchers in discussing the extension of ANOVA to MANOVA mehods. They discussed MANOVA, discriminant analysis and step down F tests as preferable to multiple univariate F tests in a multivariate experiement, asserting that only through use of these multivariate techniques can behavioral researchers derive the full benefits of multiple outcome data.

Overview of MANOVA

Before proceeding with discussion of these alternatives, a brief overview of the MANOVA technique, with an emphasis on its relationship to univariate ANOVA is presented.

The multivariate null hypothesis can be stated in a general sense as the equality of a set of dependent variables among a number of treatment populations. More specifically, the null states that the p-variate mean response vectors (dependent variable set), $\bar{\mu}_i$, are equal across k populations.

$$H_o: \bar{\mu}_1 = \bar{\mu}_2 = \ldots = \bar{\mu}_k \qquad (1)$$

The overall test of this hypothesis may be composed of k-1 non-redundant contrasts among the p-variate response vectors. This should be a familiar univariate concept as well, referring to the degrees of freedom in estimating mean differences among k populations. A set of k-1 contrasts among treatment populations is often represented by a matrix of coefficients, C, and the set of k-1 contrasts composing the null hypothesis may be represented as follows:

$$H_o: \quad C\,\mu = 0 \qquad (2)$$

where:

C = k-1 by k coefficient matrix

μ = k by p matrix of the treatment p-variate mean response vectors

0 = k-1 by p zero matrix

For example, if three populations are under study, on the basis of a single dependent variable (i.e., a univariate hypothesis), two orthogonal contrasts may be tested to fully account for the variation among treatment populations. If population 1 is compared with population 2 in contrast #1 and with population 3 in contrast #2, this set of contrasts would be represented:

$$H_o: \quad \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad (3)$$

The multivariate analogue merely replaces each population mean μ_i, with a population response vector, $\vec{\mu}_i$, containing p elements.

$$\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \vec{\mu}_1 \\ \vec{\mu}_2 \\ \vec{\mu}_3 \end{bmatrix} = \begin{bmatrix} \vec{\mu}_1 - \vec{\mu}_2 \\ \vec{\mu}_2 - \vec{\mu}_3 \end{bmatrix} = \begin{bmatrix} \vec{0} \\ \vec{0} \end{bmatrix} \qquad (4)$$

Thus, this pair of multivariate contrasts compares populations 1 and 2, and populations 1 and 3 on the entire set of dependent variables simultaneously. In linear discriminant analysis, the p dependent variables in the population vectors are all differentially weighted, forming a single composite response observation, Y_c.

$$Y_c = w_1Y_1 + w_2Y_2 + \ldots + w_pY_p \quad (5)$$

This optimal weighting procedure is determined so as to maximize the differences among the k populations or, in the case of specific group contrasts as previously illustrated, to maximize the value of the contrast.

Aside from that combination of dependent variables which maximizes the value of the contrast, the researcher may have other combinations of dependent variables in mind. Once you begin thinking of comparing treatment groups on combinations of dependent variables, the distinction of multivariate hypothesis testing from univariate hypothesis testing comes clear. There is really a second dimension to post hoc analysis of the multivariate hypothesis test - one which involves searching for contrasts among <u>dependent variables</u> as well as among treatment groups in testing for population differences.

Thus, the multivariate test can be formulated as in (2), but with the addition of another coefficient matrix, A, which contains the weights applied to the p dependent variables in testing the k-1 contrasts among treatment populations.

$$H_o: C \mu A = 0 \qquad (6)$$

Where
C = k -1 x k coefficient matrix

μ = k x p matrix of the p-variate treatment mean response vectors

A = p x a matrix of weights applied to dependent variables

0 = k-1 by a zero matrix

As previously mentioned, one set of weights which may be of interest are the discriminant function weights (w_i in (5)) which maximize the values of the contrasts specified in C. These are derived through the application of discriminant analysis and are determined on the basis of the sample data at hand. Prior to any data analysis, however, there may be combinations of variables, i.e., weighting schemes for inclusion in the A matrix, of interest to the researcher. Subsets of the p dependent variables may be combined to form new composite outcomes of interest in the test of the generic null hypothesis. Simple sums or differences among the dependent variables may also be of interest. Finally, the variables may be of interest individually in testing the group contrasts. Tests of this type are actually the familiar univariate tests, but may be considered a subset of the multivariate procedures.

To illustrate some possible combinations as described above, suppose that our previous example was comparing three populations of

National Football League quarterbacks. Population 1 consisted of quarterbacks who played in the NFL in the 1950's, population 2 were those who played in the 60's, and population 3 were those who played in the 70's. Their performances are represented by a three-variate outcome vector consisting of touchdown passes (y_1), ratio of interceptions to passes attempted (y_2), and touchdown runs (y_3).

The contrasts among the groups are as previously specified, comparing quarterbacks of the 1950's first with those of the 1960's (row 1 of the C matrix) and then with those of the 1970's (row 2 of C). Combinations of dependent variables of interest are total touchdowns ($y_1 + y_2$), the difference between number of touchdown passes and touchdown runs ($y_1 - y_2$), and each dependent variable individually. The columns of the weight matrix A represent these combinations as follows:

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 \end{bmatrix}$$

The matrix expression representing this multivariate set of hypotheses is then:

$$\underset{C}{\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}} \quad \underset{\mu}{\begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{12} & \mu_{23} \\ \mu_{31} & \mu_{32} & \mu_{33} \end{bmatrix}} \quad \underset{A}{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 \end{bmatrix}} = 0 \quad (7)$$

This approach merely sets up the questions of interest in a compact way. The statistical tests of these contrasts may be conducted in a variety of ways, however, and it is this next step that is the real focus of this paper. An important distinguishing factor among these alternatives is whether the coefficients in C and/or the weights in A have been selected *a priori*, without benefit of preliminary analysis of the data, or in a *post hoc* manner, after gaining some insight as to where population differences lie. A number of authors (Stevens, 1972; Harris, 1975; Wilkinson, 1975) have noted and demonstrated the dramatic gain in statistical power to detect alternative hypotheses achieved when some degree of *a priori* specification exists. This paper speaks more to those situations in which little is anticipated in this regard (i.e., truly post hoc analysis), but this is because it seems to be the more representative case. Admittedly, this is an unfortunate state of affairs, one which is decried by virtually every statistical or experimental design text in existence. Yet, the need for truly exploratory research is ever-present as is testified by Tukey (1977). The author will occasionally punctuate the results of post hoc tests illustrated in this paper with corresponding results had the contrasts been specified *a priori*, but the purpose of this paper is not to advance the well-worn argument of *a priori* specification vs. *post hoc* "fishing." In the end, these distinguish themselves only in their inferential probability statement - i.e., the "critical value" of the test statistic chosen - not in the use of the procedure itself. In the example analysis at the conclusion of the

paper, the author exemplifies these choices in hopefully a rational way, but sticks to the purpose of selecting the post hoc technique in line with the experimental interest.

Type I Error Rate

The multivariate null hypotheses is, as has been mentioned, two dimensional. It concerns both differences among treatment groups and differences among the dependent variables in accounting for these treatment differences. An omnibus MANOVA tests this hypothesis at a user-defined Type I error rate. If the multivariate test statistic exceeds its critical value at the selected level of Type I error probability, the researcher may proceed to any of a number of the methods covered in this paper.

Regardless of the method chosen, however, specification of the Type I error rate for the test is necessary. A number of authors have elaborated on the important principles to consider in choosing the Type I error rate for these tests. Harris (1975) presents a lucid discussion of the distinction between a priori and post hoc specification of hypotheses. In multivariate applications, this distinction relates both to specification of treatment group contrasts and dependent variable combinations. Harris includes a useful factorial diagram of the four possible combinations, one which further highlights this paper's focus on the two dimensional nature of multivariate hypothesis testing. Guidance as to the appropriate means for setting an appropriate Type I error rate (e.g., Bonferroni split per test vs. experimentwise control through simultaneous test procedures) as well as reference to the appropriate sampling

distribution (multivariate or univariate) is provided. The author has found this a useful organizational device in determining alternative data analytic methods for each of these four situations. In arriving at a method of choice, however, some additional specification is needed. That is, once the appropriate Type I error rate is set, there are often still a number of statistical tests to choose from. Again, the current paper hopes to exemplify appropriate specification of Type I error rate, but maintain its focus on demonstrating the differential applicability of the data analytic methods for different post hoc questions of interest.

OVERVIEW OF POST MANOVA TECHNIQUES

Univariate F Tests

In the author's experience, perhaps the most frequently employed post hoc method in MANOVA studies is to scan the set of dependent variables to determine which of them, if any, have evidenced statistically significant differences among the treatment levels by themselves. While this procedure is not without its advocates even in the multivariate literature (e.g., Hummel & Sligo, 1971; Finn, 1974, p. 320) users are well advised not to mistake the results of univariate tests as speaking to multivariate hypotheses. Indeed, there is no necessary relationship between the two. The intercorrelation pattern among the dependent variables dictates any correspondence between the results of multivariate and univariate tests.

The effects of the treatment on each of p dependent variables alone may very well be an important interest to the researcher, however. If so, a Bonferroni-type split of the familywise Type I error rate into p parts and screening each univariate F at the α/p level will provide an accurate test of the univariate effect. Once these have been performed, if the effect in question has more than two levels, it may be of further interest to probe the nature of the group differences on this single dependent variable. For this purpose, a wide variety of univariate multiple comparison techniques are appropriate. Once reduced to this essentially univariate problem, the user may avail himself of the excellent treatments of this diverse area (Hopkins and Chadbourn, 1967; Games, 1971; Miller, 1968). Note that the Type I

error value pertinent for these multiple comparions within the univariate hypothesis is, not the initial of the multivariate hypothesis. In choosing among multiple comparison methods, one must deal with such issues as a priori vs. post hoc testing, simple vs. complex contrasts, and experimentiwise vs. familywise vs. contrastwise error rates (Kirk, 1968).

One concern in relying on univariate F tests and multiple comparison procedures is that such a practice simply does not tap the nature of the multivariate result. Too often, the researcher responds with consternation when, for example, the results of the multivariate test are statistically significant, but none of the univariate results even approach significance. The response of the multivariate data analyst is simple (though not terribly specific) - the difference among treatment groups indicated by the multivariate test is some combination of fairly unrelated dependent variables, and none of them by themselves is significantly affected by the treatment. Taken together, however, they represent a cumulative (and nonredundant) effect of the treatment. Or the opposite extreme - some of the univariate F tests are statistically significant, but the multivariate test is not. Again, the multivariate consultant to the rescue. The problem is most likely one of redundancy. The dependent variables are probably highly intercorrelated and/or are part of a large set of dependent variables included in the analysis. The occurrence of one statistically significant result in such a large set may well be

within the realm of chance when viewed in a multivariate context. So, again, there is no necessary relationship between univariate and multivariate results.

For the researcher to employ univariate F tests exclusively in the post hoc analysis of multivariate data would be a gross disservice to the research. As only part of the post hoc effort, however, they speak to a potentially important hypothesis and are the superior method for addressing it: how is each outcome variable by itself affected by the treatment? If these are the only hypotheses of interest to the user, there is no need to perform a multivariate analysis. Simply conduct p univariate ANOVAs, each at the α/p level.

Discriminant Analysis

The application of discriminant analysis following a statistically significant MANOVA provides a multivariate alternative to univariate F tests in that it considers the intercorrelations among the dependent variables in deriving that weighting of dependent variables which maximally differentiates among the treatment groups. Linear discriminant analysis is covered in virtually every textbook on multivariate analysis, spanning theoretical treatment (Anderson, 1958) to conceptual description (Tatsuoka, 1970; Cooley & Lohnes, 1971). The nature of the weighting scheme in the discriminant function may be useful interpretively, as it defines the composite dependent variable on which at least one multivariate test is totally based (Roy's largest root criterion) and, in the absence of more than one dimension of group difference, provides useful insight into the nature of the

difference reflected in other multivariate test statistics. As stated by Harris (1975), multivariate statistical techniques are a straightforward extension of univariate techniques, with the single dependent variable represented by a composite of the original outcome variables. This composite, when dealing with the first dimension of a multivariate effect, is defined by the discriminant function. In MANOVA applications, a statistically significant multivariate result can often be meaningfully probed by pursuing comparisons among treatment group combinations on the basis of this composite variable alone. In other words, a new variable is formed as the weighted combination of original dependent variables defined by the discriminant function.

Huberty (1975) presented a thorough review of the utility of discriminant analysis and distinguished among several purposes of this versatile statistical technique. In this paper, we are concerned with only one of them, the study of maximal group differences on the basis of a linear combination of variables. When conducting a discriminant analysis, the researcher is typically interested in interpreting the nature of the composite variable on which it is based. This interpretation involves consideration of at least two indices. The standardized discriminant weights applied to each dependent variable in forming the composite, and the canonical structure coefficients which represent each variable's correlation with the resultant composite. A number of studies have included these two statistics along with several others in interpreting the discriminant function (Huberty, 1975; Wilkinson, 1975; Gabriel & Glavin, 1978). The fact

that they rarely yield the same result in identifying the important discriminators among dependent variables is not surprising in light of their fundamental difference. Discriminant weights are partial coefficients, calculated with the effects of all the variables taken into account. They reflect a variable's contribution over and above all others in the set, thus eliminating any redundancy in the discrimination. Canonical structure coefficients, however, are simply the variable-composite correlations. They are useful in that they indicate which individual variables are most like the composite criterion, but do not partial out the interrelationships of the other variables thereby identifying possibly redundant discriminators as important. Taken together, these two provide valuable information toward the interpretation of the composite.

Particularly when the composite is an interpretively meaningful combination of variables (Tatsuoka, 1973), multiple comparison techniques may be pursued further using this composite score as the dependent variable. The fact that it is a combination of variables derived on sample data, however, requires that a more complex sampling distribution be taken into account when significance tests are performed. Extensions of the familiar univariate multiple comparison techniques which apply to treatment level comparisons based on a composite of original variables are presented by Bird (1975) and Harris (1975). The contrast specification is the same, but the critical value for the test statistic differs from that used in strictly univariate multiple comparison techniques.

Step Down F Tests

Univariate F tests consider each dependent variable singly without regard to its intercorrelation with any of the others, while discriminant analysis considers each dependent variable's intercorrelation with the entire set. Step down F tests may be considered as falling between these two extremes. Differences among treatment levels on a given dependent variable are tested after those differences evidenced on previous dependent variables are statistically partialled out. Thus, as the analysis proceeds through the set of dependent variables, treatment differences are sought in the residuals of successive variables. The step down F test of the jth variable of the set tests for differences among the treatments which have not already been accounted for by the previous j-1 variables. It is equivalent to an analysis of covariance, treating all previous variables as covariates to the one under consideration. Consequently, its set of results on p dependent variables are dependent upon the order in which they are analyzed. This is a limiting factor in its meaningful application. When such an ordering is meaningful, however, it affords the researcher an orthogonal partition of the p-variate sum of the between group sums of squares. The familywise set for the multivariate hypothesis can be partitioned using the following well known relation appropriate to a set of orthogonal elements.

$$\alpha = 1 - (1 - \alpha_1)(1 - \alpha_2) \ldots (1 - \alpha_p)$$

There is no requirement that the α_i be equal. Bock and Haggard (1968) suggest the experimenter choose the order and assign the Type I error rate according to a priority of interest on the dependent variables. Those variables of greater interest should be placed higher up in the ordering and assigned more liberal Type I error rates.

Some natural orderings of dependent variables may exist in social science research. If measures are taken over time and considered a response vector in the study (e.g., repeated trials in an experiment, or observations across months or years on a longitudal study) their chronology suggest an ordering, and step down F tests would investigate treatment group differences not already observed at previous points in time. While other data analytic methods may be preferable given other characteristics of this type of research design (multiple profile analysis, repeated measures ANOVA, time series analysis), step down F tests may perform a meaningful function among such variables. Other applications which suggest themselves are those dealing with hierarchical constructs, such as Bloom's (1956) taxonomy of cognitive educational objectives. Finally, if the area of application has a typical set of outcome variables on which the effects of treatment are measured, and the interest is in determining if a new criteria construct evidences any effect not already manifested in other criteria, the sequential nature of the step down F procedure fits this purpose nicely. Like adding a new predictor variable to an established regression equation last, analyzing the new dependent variable last on the outcome vector will provide this information.

Profile Analysis

When dependent variables are measured on the same scale and form a conceptually meaningful "profile" of the respondent, certain patterns across this profile may be of interest. Operationally, this translates to particular contrasts among the dependent variables. Consider an example in which scores on three psychological scales are being compared across two treatment groups. Sample results are shown in Figure 1. Since the three measures may be said to describe a psychological "profile" of each group, at least three questions regarding this profile consistently emerge. These questions are represented by hypotheses of Parallelism, Levels, and Flatness in a profile analysis of such data, and are explicated below.

Parallelism: Do the two treatment group profiles show the same pattern or shape across the three scales?

Levels: Are the overall levels across the three scales in the profiles the same?

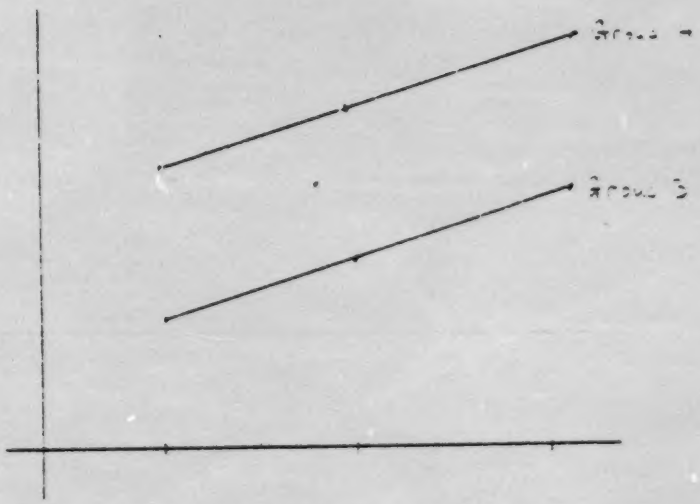
Flatness: Are the profiles flat? That is, are there no differences across scales within a given profile?

Figure 1 represents a case in which the hypothesis of Parallelism would not be rejected - the two groups show the same pattern, a steady increase, across the three scales. The Levels hypothesis would be

Figure 1

Comparison of Psychological Profiles

of Two Treatment Groups



rejected, however. Treatment group A is at a considerably higher level than B on all three scales. The Flatness hypothesis would also be rejected. There are differences across measures within the two profiles, the consistent increase as noted previously.

This type of data has frequently been analyzed using a univariate repeated measures design, where the three psychological scales would constitute a repeated measures factor in the design. Although this is a different analytic approach possessing restrictive assumptions rarely satisfied in social science research, the familiar hypotheses of the Groups by Measures repeated measures design are identical to those of profile analysis. The main effect of Groups tests the equality of the treatment groups across all measures. This is precisely what is tested by the Levels hypothesis in profile analysis. The main effect of Measures looks for differences among the measures - the same question addressed by the Flatness hypothesis. Finally, the Group by Measures interaction is identically a test of Parallelism of profiles. That is, are the differences among groups the same for all measures?

Repeated Measures ANOVA	=	Profile Analysis
Treatment Group Main Effect	=	Levels
Measures Main Effect	=	Flatness
Group by Measures Interaction	=	Parallelism

Simultaneous Test Procedures

The post MANOVA techniques already discussed deal primarily with the dependent variable dimension of the experimental design. When concerned with probing the treatment group dimension for differences, the well known univariate multiple comparison procedures (Games, 1971) can be used only when single dependent variables are of interest. Extensions of some of these methods (the Tukey, Sheffe and Bonferroni-t techniques) to the multivariate case presented by Bird (1975) are applicable when the analysis is partially or fully planned. When the analysis is entirely post hoc, that is, when the selection of treatment groups and dependent variables for comparison is done only after viewing the results of the omnibus MANOVA, the researcher may employ simultaneous test procedures (STP) based upon any of the commonly employed multivariate test statistics (Gabriel, 1968; Gabriel, 1969). Often described as multivariate analogues to the univariate Scheffe' simultaneous test procedure, these usually produce conservative results and are rarely recommended. This conservativism is a necessary consequence of the wide variety of contrasts the methods may address within a single experimentwise expression of Type I error rate. Unlike previous methods, the two dimensions of differences discussed earlier may be tested in combination using these techniques. Treatment group contrasts may be tested on the basis of dependent variable combinations as shown in (6) and (7), and both may be selected after viewing the data. The researcher pays for this flexibility, however, with the lack of sensitivity of the tests.

An Example

Results from a study conducted to determine the mediating effects of various types of causal attribution of success or failure in a task in a learned helplessness paradigm will serve as an extensive example for the set of multivariate methods discussed in this paper. Briefly, the design of the study included two factors. Eighty subjects were randomly assigned to either a solvable or unsolvable initial anagram task and their performance on that task was subsequently attributed to one of four possible causes: their inherent ability in performing such tasks, their effort in attempting the task, the difficulty of the task itself, or luck. The particular cause attributed was also randomly assigned to subjects. The 2 x 4 design (n = 10 per cell) is shown in Figure 2.

Dependent variables measured in the study were of essentially two types: measures of actual performance on a subsequent task assigned (which was solvable for all subjects) and measures of affect following the subjects completing the task. The three measures of performance and three measures of affect are also listed in Figure 2.

Figure 2

Experimental Design of Example Study

	Ability	Effort	Task Difficulty	Luck
Solvable	(n = 10)			
Unsolvable				

Dependent Variables:

Performance	Affect
Number of Trials	Anxiety
Number of Failures	Depression
Response Latency	Hostility

Before reviewing the results of the MANOVA carried out in these data, some preliminary analytic strategies are worth considering. Prior thinking on the treatment factors in the design led the experimenter to stipulate two complex contrasts among the levels of Attribution. These would be tested both across levels of the Solvability factor (i.e., as in the Attribution main effect) and in contrast with the two levels of Solvability (i.e., the S x A interaction). The first contrast combines the Ability and Effort attributions and compares them to Task Difficulty and Luck. This was thought to represent a possible Locus of Control dimension among the four types of causal attribution included in this study - ability and effort being internal characteristics of the subject while task difficulty and luck are external to his control. The second contrast combines Ability and Task Difficulty and compares them with Effort and Luck. This was thought to compare the attributions which represented stable characteristics of the subject and task (ability and task difficulty) with unstable characteristics (task difficulty and luck). In the remainder of this section, these contrasts will be referred to as Internal vs. External and Stable vs. Unstable, respectively.

Framing these contrasts in the matrix expression of the null hypothesis as previously described involves specification of the contrast matrix C as follows:

H_o: $C \cdot A = 0$

$$C = \begin{bmatrix} 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}$$

The dependent variables were selected for study with the notion that they represented two quite separate outcomes of the experiment - the subject's performance on the experimental task and his affective reaction following the pre-treatment anagram problem. A look at the pooled within cell correlation matrix in Table 2 begins to lend empirical support to this notion. Variables within these two groups evidence high correlations, particularly the three performance measures - all in the neighborhood of .80 and higher. The correlations across the two sets show little relationship.

TABLE 2

Within Group Correlation Matrix*

	Trials	Failures	Latency	Anxiety	Depres-sion	Hostil-ity
Trials	40.73					
Failures	.79	14.46				
Latency	.87	.97	459.51			
Anxiety	.11	.02	.07	4.77		
Depression	.00	.02	- .01	- .37	3.33	
Hostility	- .31	- .11	- .17	- .30	.50	3.56

* Diagonal elements are pooled within group variances.

TABLE 3

PRINCIPAL COMPONENTS OF DEPENDENT VARIABLES

	1	2	3	4	5	6
Trials	-.93	-.09	-.02	-.15	-.33	.03
Failures	-.93	-.24	-.02	-.16	.21	.08
Latency	-.97	-.19	-.04	-0.10	.07	-.11
Anxiety	-.19	.66	.71	.11	.02	.00
Depression	.13	-.82	.21	.50	.08	.00
Hostility	.37	-.71	.41	-.42	-.12	.00

Component	λ	% Variation	Cummulative %
1	2.85	47.46	47.46
2	1.74	28.92	76.38
3	.72	12.04	88.42
4	.50	8.32	96.74
5	.18	2.95	99.69
6	.02	.31	100.00

A principal components analysis was conducted and it was found that the first two componenets extracted from the set of six variables accounted for over 75% of the variance among them. Table 3 contains the component loadings of the six variables and the reader can see the obvious separation of performance variables from affect variables. It was decided on this basis to perform both a MANOVA on the entire set of six outcome variables and separate MANOVAs on the two subsets. Although there is great homogeneity among variables within the two subsets in the same treatment condition (i.e., Table 2 correlations are calculated within cells and pooled across cells much like within cell variances are in ANOVA), this does not guarantee that they will be affected in the same ways by a treatment manipulation. The MANOVA on all six will lend insight into how performance and affect are mutually affected by treatment manipulation. There are at least two possibilities of interest.

1. Even though they are virtually uncorrelated <u>within</u> a treatment condition, the differences across treatments could reveal common effects. This would be evidenced by only one being heavily influential in the discriminant function (the other being redundant), and the step down F's of the variables ordered last being greatly reduced from their univariate F statistics.

2. They could operate quite separately and potently in their own rights. If only one discriminant function were derivable (i.e., a single degree of freedom test), representatives from both subsets would be heavily weighted. If more than one were derivable, perhaps the variables would load heavily on separate discriminant functions. Step down F values of the variables ordered later in the set of six would remain nearly as high as their univariate F's.

By conducting a MANOVA on the two subsets separately, the experimenter gains insight into the between-group redundancy of individual variables within that common construct (performance or affect). It will be useful to know if a great deal of "excess baggage" is included in these subsets in terms of the treatment manipulation's effect on them.

The author is aware that this was to be a paper on post hoc analysis of MANOVA results. Interpretation of the interrelationships among the dependent variables will aid in planning the most informative multivariate approach to his data, however. It is important to point out that the foregoing correlational analyses are performed within treatment combinations and are independent of any mean differences among the groups. Therefore, any selection of variables on these bases can be rightly considered a priori, and the experimenter is not yet in the arena of multivariate "data snooping" for group mean differences which introduces the array of multivariate simultaneous test procedures (Gabriel, 1968; Morrison, 1976) which

have been demonstrated to be woefully conservative and hopefully a last resort for the multivariate researcher (e.g., Stevens, 1973; Bird, 1975). In terms of Type I error specification, then, the experimenter would be justified in employing a split of his experimentwise Type I error rate according to the Bonferroni inequality previously discussed.

Table 4 contains the results of the overall six variate MANOVA of the 2 x 4 experimental design. Statistics presented include univariate Fs, step down Fs, and raw and standardized discriminant function coefficients (b_i) along with canonical structure coefficients (r_c) for the leading discriminant function for all variables and effects. The multivariate F statistic presented merits some explanation, as has been lucidly pointed out by Olson (1976). It is derived from Wilk's (1932) likelihood ratio according to the function developed by Rao (1952). Commonly termed "Rao's approximation" to the F distribution, it affords ready access to conventional F sampling distribution and is, in fact, exact in some instances of p (number of dependent variables) and k (number of treatment groups). Many of the tests performed in this sample analysis will be exact and those that are not are of the experimental design conditions where Rao's approximation is highly accurate.

The multivariate tests indicate that there is a statistically significant main effect of Solvability ($F = 4.15$, $p < .001$) and interaction effect of Solvability and Attribution ($F = 1.76$, $p < .032$). Interpretively, all that can be said at this point is that

subjects who received unsolvable problems differed from those who received solvable problems in some way or ways reflected in the six variate outcome set. Since there are only two groups being compared, the post hoc effort can concentrate on the dependent variable combination which best account for the multivariate difference. A look at the univariate F results shows only two variables affected individually by the treatment---affect ratings of depression ($F = 20.55$, $p < .0001$) and hostility ($F = 5.35$, $p < .024$). None of the performance variables by themselves even suggest significant differences. The step down F results further modify these conclusions by identifying the redundancy of the effects on hostility to those of depression. The step down F for hostility ($F = .19$, $p < .666$) is far smaller than its univariate F, indicating its effects were almost totally predictable by those observed on the self rating of depression. This is not to say that depression ratings were affected and hostility ratings were not. It merely says there is <u>one</u> effect in evidence here, not two, and it is one that is shared by the common variance of hostility and depression. The standardized discriminant weights show a large value applied to depression ($b_1 = -1.02$) and a near-zero weight applied to hostility ($b_1 = .13$), reflecting its redundancy to the discrimination. This redundancy is not taken into account by the canonical structure coefficients, however, as the correlations of depression ($r = -.90$) and hostility ($r = -.82$) with the composite are both extremely high and represent the only variables of the six which are strongly related to the composite. The interpretation seems clear that the multivariate Solvability main effect is based upon the affect common to depression and hostility.

TABLE 4

2 x 4 MANOVA on Both Performance and Affect Variables

	df	MS	Univ. F	p	Step Down F	p	raw b_i	std. b_i	r_c
Solvability	6		4.15	.001					
Trials	1	43.51	1.07	.305	1.07	.305	.16	1.00	.23
Failures	1	6.61	.46	.501	.05	.821	.44	1.65	.16
Latency	1	121.40	.26	.608	2.48	.120	-.11	-2.35	.15
Anxiety	1	6.05	1.27	.264	1.49	.225	-.04	- .08	.21
Depression	1	68.45	20.55	.0001	18.55	.0001	-.56	-1.02	-.90
Hostility	1	19.01	5.35	.024	.19	.666	.07	.13	-.82
Attribution	18		1.12	.337					
Trials	3	77.15	1.89	.138	1.89	.138	.33	2.12	
Failures	3	14.31	.99	.403	2.77	.043	-.07	- .27	
Latency	3	423.37	.92	.435	.44	.726	-.06	-1.35	
Anxiety	3	.85	.18	.911	.16	.925	-.04	- .10	
Depression	3	1.02	.31	.822	.22	.881	-.28	- .50	
Hostility	3	.71	.20	.896	1.45	.236	.34	.64	
S & A	18		1.76	.032					
Trials	3	66.18	1.62	.191	1.62	.191	-.19	-1.18	-.22
Failures	3	4.78	.33	.803	1.82	.151	.54	2.05	-.10
Latency	3	244.19	.53	.662	.76	.519	-.06	-1.27	-.14
Anxiety	3	3.12	.65	.584	.72	.542	-.04	- .09	-.34
Depression	3	12.08	3.63	.017	3.09	.033	.05	.09	.87
Hostility	3	5.21	1.47	.231	2.51	.066	-.43	- .82	.67
Within									
Trials	72	40.73							
Failures	72	14.46							
Latency	72	459.51							
Anxiety	72	4.77							
Depression	72	3.33							
Hostility	72	3.56							

The relatively high standardized discriminant weights in the three performance variables are interesting. None of their univariate F tests approach significance, their step down F results are not suggestive, and none correlate highly with the discriminant composite (r_c). The magnitude of their weights may represent a suppressor effect (Darlington, 1968), intercorrelating highly with that small portion of the variance of the other more discriminating variables which is unrelated to group differences, thereby making them more efficient discriminators themselves. Further insight into this may be gained by comparing the results of the six variate MANOVA with those of the three variate MANOVAs on performance and affect variables separately.

TABLE 5

Cell Means of Performance and Affect Variables

	Ability		Effort		Task Difficulty		Luck	
	Perf.	Affect	Perf.	Affect	Perf.	Affect	Perf.	Affect
Solvable	13.50	4.90	18.70	4.60	12.90	5.40	14.80	5.30
	4.00	8.10	6.60	8.90	5.30	7.10	4.50	7.90
	32.26	8.20	49.35	9.30	37.40	7.90	36.61	8.40
	Perf.	Affect	Perf.	Affect	Perf.	Affect	Perf.	Affect
Unsolvable	20.10	6.10	16.70	5.80	13.60	5.60	15.40	4.90
	5.30	5.50	6.60	5.30	4.80	7.00	6.00	8.10
	42.97	7.60	44.13	7.10	36.475	8.10	40.90	8.20

TABLE 6

2 x 4 MANOVA on Performance Variables

	df	MS	Univ. F	p	Step Down F	p	raw b_i	std. b_i	r_c
Solvability	3		1.20	.315					
Trials	1	43.51	1.07	.305	1.07	.305	-.32	-2.06	
Failures	1	6.61	.46	.501	.50	.821	-.86	-3.27	
Latency	1	121.40	.26	.609	2.48	.120	.22	4.71	
Attribution	9		1.67	.100					
Trials	3	77.15	1.89	.138	1.89	.138	.31	2.00	.51
Failures	3	14.31	.99	.403	2.77	.048	-.05	- .21	-.08
Latency	3	423.37	.92	.435	.44	.726	-.07	-1.46	.04
S x A	9		1.38	.198					
Trials	3	66.18	1.62	.191	1.62	.191	.17	1.07	
Failures	3	4.78	.33	.803	1.82	.151	-.64	-2.41	
Latency	3	244.19	.53	.662	.76	.519	.08	1.69	

The interpretation of the S x A interaction cannot merely concentrate on the dependent variables. There are as many as three nonredundant sources of group differences (i.e., three degrees of freedom) to be considered as well as probing into the nature of these effects among the dependent variables. Experimental interest has been specified in two of those sources already - the Internal vs. External and the Stable vs. Unstable contrasts. Discussion of the S x A effect will be postponed until these are tested individually.

The separate MANOVAs performed on performance and affect variables are presented in Tables 6 and 8 respectively. The performance variables evidence no main effect of Solvability ($F = 1.20$, $p < .315$) or overall interaction of Solvability and Attribution ($F = 1.39$, $p < .196$), lending some the interpretation of the six variate MANOVA of these effects as concentrated on the affect variables. Since there are three degrees of freedom operating in the S x A interaction, however, and since two contrasts of the levels of Attribution were of interest a priori, there is no need to cease the analysis of the S x A interaction. The overall test warns that, if there is a significant group contrast operating, its single degree of freedom effects are probably all that are operating among the three degrees of freedom and are being watered down by representation as three degrees of freedom in the overall test. These two contrasts and post hoc tests will be presented later.

The main effect of Attribution may be considered statistically significant here ($F = 1.67$, $p < .10$). A look at the univariate F results does not provide the key to the interpretation of this main

effect. Only Trials approaches consideration for interpretation ($F = 1.89$, $p < .138$), but given the multitude of tests performed and the fineness of the partition of Type I error rate to this point, this would probably exceed even liberal standards in this regard. At least this should tell the researcher to look elsewhere to determine the differences among the four levels of Attribution. Again, more specific tests of the two Attribution contrasts will be presented later.

The step down F results provide the most interesting interpretation of the Attribution main effect. The number of Failures, while not at all significantly different across the attribution levels by itself (univariate $F = .99$, $p < .403$), does evidence significant differences when the effects on the number of Trials have been partialled out (step down $F = 2.77$, $p < .048$). That is, if you consider the four groups as equal in the average number of trials required to reach criterion performance levels, they differ significantly on the number of failures they experienced in getting there. To further determine how the four groups differ on this residualized failure variable, univariate multiple comparison procedures on the residual means may be employed. These residual means are presented in Table 7. Applying Tukey's HSD method at a familywise alpha of .05 showed one significant difference. The Ability attribution group evidenced fewer failures than did the Task Difficulty group. Complex contrasts will be presented later.

TABLE 7

Attribution Group Means on Trials, Failures and Residualized Failures

	Trials	Failures	Failures/Trials*
Ability	16.80	4.65	4.14
Effort	17.70	6.60	5.66
Task Difficulty	13.25	5.05	6.20
Luck	15.10	5.25	5.54

*Residualized error mean square = 5.49

The MANOVA on the affect variables is presented in Table 8. A highly significant solvability main effect is evidenced ($F = 6.77$, $p < .0005$), one that is stronger than that of the six-variate MANOVA ($F = 4.15$, $p < .001$). This illustrates the role of "excess baggage" played by the performance variables in the overall analysis. Again, the post hoc effort need concern itself only with uncovering the effects on the dependent variables as this is a single degree of freedom treatment group effect. The Attribution main effect is virtually zero ($F = .285$, $p < .978$). Interestingly, this indicates the "excess baggage" role the affect variables played in the six variate analysis of the Attribution main effect. The overall S x A interaction is moderately significant at best ($F = 1.54$, $p < .138$), although what might be considered a highly significant univariate result ($F = 3.62$, $p < .017$) is suffering from inclusion in an otherwise unaffected package. Any post hoc interpretation of this effect will be reserved for the test of the two contrasts.

An immediate interpretation of the Solvability main effect is available, however. Table 5 contains the cell means on all six variates. Subjects who wrestled with unsolvable problems rate themselves as being significantly more depressed about their performance than those who dealt with solvable problems ($F = 10.55$, $p < .0001$). This result is not explainable through the effects on Anxiety, as the step down F for the Depression scale remains highly significant (step down $F = 19.01$, $p < .0001$). In combination with both Anxiety and Hostility, its standardized weighting on the

TABLE 8

Results of 2 x 4 MANOVA on Affect Variables

	df	MS	Univ. F	p	Step Down F	p	raw b_i	std. b_i	r_c
Solvability	3		6.77	.0005					
Anxiety	1	6.05	1.27	.264	1.27	.264	-.06	- .14	.26
Depression	1	68.45	20.55	.0001	19.01	.0001	-.56	-1.02	-.99
Hostility	1	19.01	5.35	.024	.01	.914	-.02	- .03	-.51
Attribution	9		.285	.978					
Anxiety	3	.85	.18	.911	.18	.911	.138	.302	
Depression	3	1.02	.31	.821	.21	.889	-.515	-.939	
Hostility	3	.71	.20	.896	.47	.701	.493	.930	
S x A	9		1.54	.138					
Anxiety	3	3.12	.65	.584	.65	.584	-.01	-.03	.35
Depression	3	12.08	3.62	.017	3.16	.030	-.56	-1.01	-1.00
Hostility	3	5.21	1.47	.231	.91	.443	.01	.01	-.49

discriminant function shows it to be the only important contributor of the three to the multivariate difference. Its canonical structure coefficient is a correlation of .99. Thus, as in the six variate MANOVA, it seems obvious that the affective difference between solvable and unsolvable conditions can be virtually fully accounted for by the self rating of depression along with its commonality with hostility.

Attribution Contrasts

The test of the two contrasts of the Attribution factor may be conducted with respect to both the Attribution main effect and the Solvability by Attribution interaction. All six dependent variables may be included to determine if there is any interplay in the effects of the treatment on performance and affect variables. Recall that, on the basis of the overall MANOVA, there was not. The Solvability main effect was evidenced solely in the affect measures. The overall tests of the Attribution main effect and the S x A interaction do not indicate that these two subsets of outcomes are supplementing each other in producing a more dramatic multivariate difference than is observed within each set separately. On the contrary, the subsets have been viewed as "excess baggage" in one or more overall effects. The contrasts of the Attribution levels, however, could possibly tell a different story, and therefore we will first look at their tests with respect to all six dependent variables.

Tables 9 and 10 present the results of the two contrasts tested on the basis of all six dependent variables for the Attribution main effect and S x A interaction, respectively. In terms of the

Attribution main effect, the Internal vs. External comparison may be interpreted as statistically significant ($F = 2.28$, $p < .046$), while the Stable vs. Unstable yields no such indication ($F = .63$, $p < .710$). This is enlightening, in that the overall test of the main effect was not near enough to significance to warrant even cautious interpretation ($F = 1.12$, $p < .337$, c.f. Table 4). It appears that a statistically reliable Internal vs. External difference was indeed buried in the overall three degree of freedom effect. The interpretation of this Attribution contrast in terms of dependent variables is similar to that of the overall Attribution main effect (c.f. Table 6). Only Trials has a univariate F which approaches significance. The step down F of Failures again communicates the difference here. The Internal Attribution condition evidenced significantly fewer failures when partialling out the differences in number of Trials between it and the External Attribution conditions.

The S x A interaction presented in Table 10 shows a significant Stable vs. Unstable comparison ($F = 2.78$, $p < .018$) across the two levels of Solvability. Relevant variables in this effect appear to be affective, with the step down result for Failures again being much more significant than its univariate alone. Interpretation of invariate F tests and discriminant analysis, however, point to depression and hostility as reflecting the differences in this contrast.

TABLE 9

Results of Attribution Main Effect Contrasts
on Both Performance and Affect Variables

	df	MS	Univ. F	p	Step Down F	p	raw b_i	std. b_i	r_c
Internal/ External	6		2.28	.046					
Trials	1	126.03	3.09	.08	3.09	.083	.34	2.17	.51
Failures	1	1.60	.11	.74	7.50	.008	-.06	-.22	-.03
Latency	1	4.59	.01	.92	.75	.390	-.07	-1.57	.16
Anxiety	1	.00	.00	1.00	.09	.759	-.05	-.11	.01
Depression	1	.63	.19	.67	.31	.577	-.23	-.42	-.14
Hostility	1	.10	.03	.87	1.82	.181	.27	.51	-.29
Stable/ Unstable	6		.63	.710					
Trials	1	95.41	2.34	.310	2.34	.13			
Failures	1	40.83	2.82	.097	.56	.46			
Latancy	1	1194.17	2.60	.111	.16	.69			
Anxiety	1	1.20	.25	.618	.26	.61			
Depression	1	.41	.12	.727	.01	.93			
Hostility	1	.83	.23	.630	.57	.45			

TABLE 10

Results of S x A Interaction Contrasts
on Both Performance and Affect Variables

	df	MS	Univ. F	p	Step Down F	p	raw b_i	std. b_i	r_c
Internal/									
External	6		1.53	.18					
Trials	1	87.03	2.14	.148	2.14	.148	.25	1.59	
Failures	1	8.10	.56	.457	.92	.517	.23	.88	
Latency	1	338.43	.74	.394	.42	.516	-.09	-1.86	
Anxiety	1	2.50	.52	.472	.36	.553	-.01	-.02	
Depression	1	15.63	4.69	.034	4.21	.044	-.55	-1.00	
Hostility	1	1.60	.45	.505	1.49	.227	.29	.55	
Stable/									
Unstable	6		2.78	.013					
Trials	1	106.41	2.61	.111	2.61	.111	-.11	-.70	-.31
Failures	1	.53	.04	.848	3.00	.088	.63	2.40	.04
Latency	1	340.88	.74	.392	1.77	.188	-.09	-2.02	-.21
Anxiety	1	.83	.17	.677	.91	.344	-.01	-.02	.09
Depression	1	16.88	5.07	.028	4.34	.041	-.16	-.29	-.56
Hostility	1	13.33	3.75	.057	3.14	.081	-.32	-.61	-.77

Since there is so little indication of any supplementary contribution to the multivariate test on the part of the performance and affective subsets, discussion of further post hoc methods will be conducted in the context of the tests of the contrasts done separately for the performance and affective variables.

Table 11 contains the results of the Internal vs. External and Stable vs. Unstable comparisons on the performance variables. When testing across levels of the Solvability factor (i.e., the Attribution main effect), the Internal vs. External comparison yields a statistically significant difference ($F = 3.86$, $p < .013$). When contrasted between levels of the Solvability dimension (i.e., the S x A interaction) the Stable vs. Unstable comparison results in a perhaps marginally significant effect ($F = 2.50$, $p < .066$). Recall that these a priori contrasts represent a division of Type I error rate of the overall group effect tested, and the chosen for the overall effect was itself already split between performance and affective MANOVA. Thus, the Type I error rate used at this level of testing should be, according to the Bonferroni procedure, approximately one-fourth of that selected for the overall MANOVA represented in Table 4. An interpretation of a difference with a p-value of .066 at this contrast-testing level, for example, is thus indicative of a willingness to test the six-variate S x A interaction with a .264 Type I error probability--a criterion few researchers would accept. The interpretation of the Internal vs. External contrast mentioned above, however, reflects willingness to test the six variate Attribution main effect with a Type I error rate of .052 - a considerably more palatable prospect.

TABLE 11

Results of Tests of Attibution
and S x A Contrasts on Performance Variables

	df	MS	Univ. F	p	Step Down F	p	raw b_i	std. b_i	c
Attribution									
Internal/ External	3		3.86	.013					
Trials	1	126.03	3.09	.083	3.09	.083	.32	2.01	.58
Failures	1	1.60	.11	.740	7.50	.008	-.05	-.20	-.03
Latency	1	4.59	.01	.921	.75	.390	-.07	-1.53	.09
Stable/ Unstable	3		1.00	.396					
Trials	1	95.40	2.34	.130	2.34	.130	-.11	-.69	
Failures	1	40.83	2.82	.097	.56	.455	-.44	-1.68	
Latency	1	1194.17	2.60	.111	.16	.692	.06	1.32	
S x A									
Internal/ External	3		.98	.406					
Trials	1	87.03	2.14	.148	2.14	.148	.31	1.96	
Failures	1	8.10	.56	.457	.42	.517	.26	.98	
Latency	1	338.43	.74	.394	.43	.516	-.10	-2.13	
Stable/ Unstable	3		2.50	.066					
Trials	1	106.41	2.61	.111	2.61	.111	.09	.57	.51
Failures	1	.53	.04	.848	3.00	.088	-.83	-3.14	.00
Latency	1	340.88	.74	.392	1.77	.188	.13	2.84	.23

Further analysis of the Internal vs. External attribution contrast would again focus on the step down F test of the number of Failures measured after partialling out group differences in the number of trials to criterion ($F = 7.50$, $p < .008$). It is clear that this interesting relationship between Trials and Failures exists only in addressing this particular contrast of the Attribution groups. The mediating effects of number of Trials is not present in the Stable vs. Unstable comparison, where the step down test ($F = .56$, $p < .455$) of Failures is quite a bit smaller than its univariate ($F = 2.82$, $p < .097$).

The tests of these attribution contrasts on the affective variables, do not result in statistically significant findings in the multivariate case in light of the previous discussion of Type I error rate appropriate for the tests of these contrasts.

Profile Analysis

The three self ratings on the anxiety, depression and hostility scales were included in the study to represent an affective outcome package. When a conceptually related set of variables measured on a comparable scale (a 1 to 11 bipolar continuum in this study) such as these are employed, the technique of profile analysis may provide interesting comparisons among treatment groups on the profiles of their response vectors.

TABLE 12

Results of Test of Attribution and S x A
Contrasts on Affect Variables

	df	MS	Univ. F	p	Step Down F	p	raw b_i	std. b_i	r_c
Attribution									
Internal/ External	3		.07	.976					
Anxiety	1	.00	.00	1.000	.00	1.00	-.175	-.38	
Depression	1	.63	.19	.666	.21	.645	-.608	-1.11	
Hostility	1	.10	.03	.867	.00	.974	.044	.08	
Stable/ Unstable	3		.12	.947					
Anxiety	1	1.20	.25	.618	.25	.618	.287	.627	
Depression	1	.41	.12	.727	.03	.860	-.028	-.052	
Hostility	1	.83	.23	.630	.09	.768	-.305	-.576	
S x A									
Internal/ External	3		1.60	.198					
Anxiety	1	2.50	.52	.472	.52	.472	-.01	-.010	
Depression	1	15.63	4.69	.034	4.09	.047	-.60	-1.100	
Hostility	1	1.60	.45	.505	.21	.650	.13	.25	
Stable/ Unstable	3		2.03	.117					
Anxiety	1	.83	.17	.677	.17	.677	-.12	-.25	-.16
Depression	1	16.88	5.07	.028	5.00	.029	-.41	-.76	-.89
Hostility	1	13.33	3.75	.057	.916	.342	-.25	-.47	-.77

TABLE 13

Profile Analysis of Affect Variables*

	df	MS	Univ. F	p	Step Down F	p	raw b_i	std. b_i
Flatness	2		28.43	.0001				
S1	1	245.00	22.21	.0001	22.21	.0001	.26	.88
S2	1	63.01	18.38	.0001	26.72	.0001	.44	.82
Parallelism Solvability	2		5.86	.004				
S1	1	115.20	10.44	.002	10.44	.002	.25	.83
S2	1	15.31	4.47	.038	1.26	.267	-.20	-.37
Attribution	6		.40	.875				
S1	3	3.43	.31	.817	.31	.817	-.09	-.31
S2	3	2.21	.65	.589	.50	.682	.47	.86
S x A	6		1.48	.188				
S1	3	25.03	2.27	.088	2.27	.088	.23	.76
S2	3	5.05	1.47	.230	.75	.525	-.25	.46
Levels								
Solvability	1	103.51	10.66	.002				
Attribution	3	.71	.07	.974				
S x A	3	20.65	2.13	.104				
Internal/ External	1	13.23	1.36	.247				
Stable/ Unstable	1	46.88	4.83	.031				

*S_1 = (Anxiety) - (Depression)

S_2 = (Depression) - (Hostility)

Table 13 presents the results of the profile analysis conducted on the three-variate affective response vector. The Flatness hypothesis is soundly rejected ($F = 28.43$, $p < .0001$), indicating there are differences among the three rating scales across all design factors. This is a very general result and one which is sufficiently mediated by subsequent tests in the profile analysis to not merit further interpretation here.

The test of Levels was highly significant when comparing the profiles of the two Solvability conditions ($F = 103.51$, $p < .002$) and perhaps marginally so for the S x A interaction ($F = 2.13$, $p < .104$). The two attribution contrasts were also tested here, yielding a statistically significant difference in levels for the Stable vs. Unstable comparison across Solvability conditions ($F = 4.83$, $p < .031$).

Figure 3 displays the Solvability x Stability interaction. The difference in overall Level of Affect is depicted in the graph of the interaction. In the solvable condition, those subjects whose performance was attributed to unstable characteristics (luck and effort) indicated negative affect more strongly than did those who were given stable attribution. Those who had dealt with unsolvable problems, however, reported more negative affect when their performance was attributed to stable characteristics (ability and task difficulty). Also in the figure is an indication of the mean Levels of Affect for the Solvability main effect, though this main effect is clearly mediated by the interaction.

In the test of the Parallelism hypothesis, only the Solvability main effect showed significantly nc parallel Affect profiles ($F = 5.86$, $p < .004$). Figure 4 presents the profiles over three ratings of affect for the two solvability conditions. The two groups are plainly different and to a similar degree on their ratings of depression and hostility, the unsolvable group reporting more negative affect on those two. The consistency of this difference on affect is disrupted by the anxiety measure, however, where the groups are essentially at the same level. Hence, their affective profiles do not show the same pattern across the three measures, and the hypothesis of parallelism is rejected.

FIGURE 3

Solvability x Stability Interaction
Difference in General Level of Affect

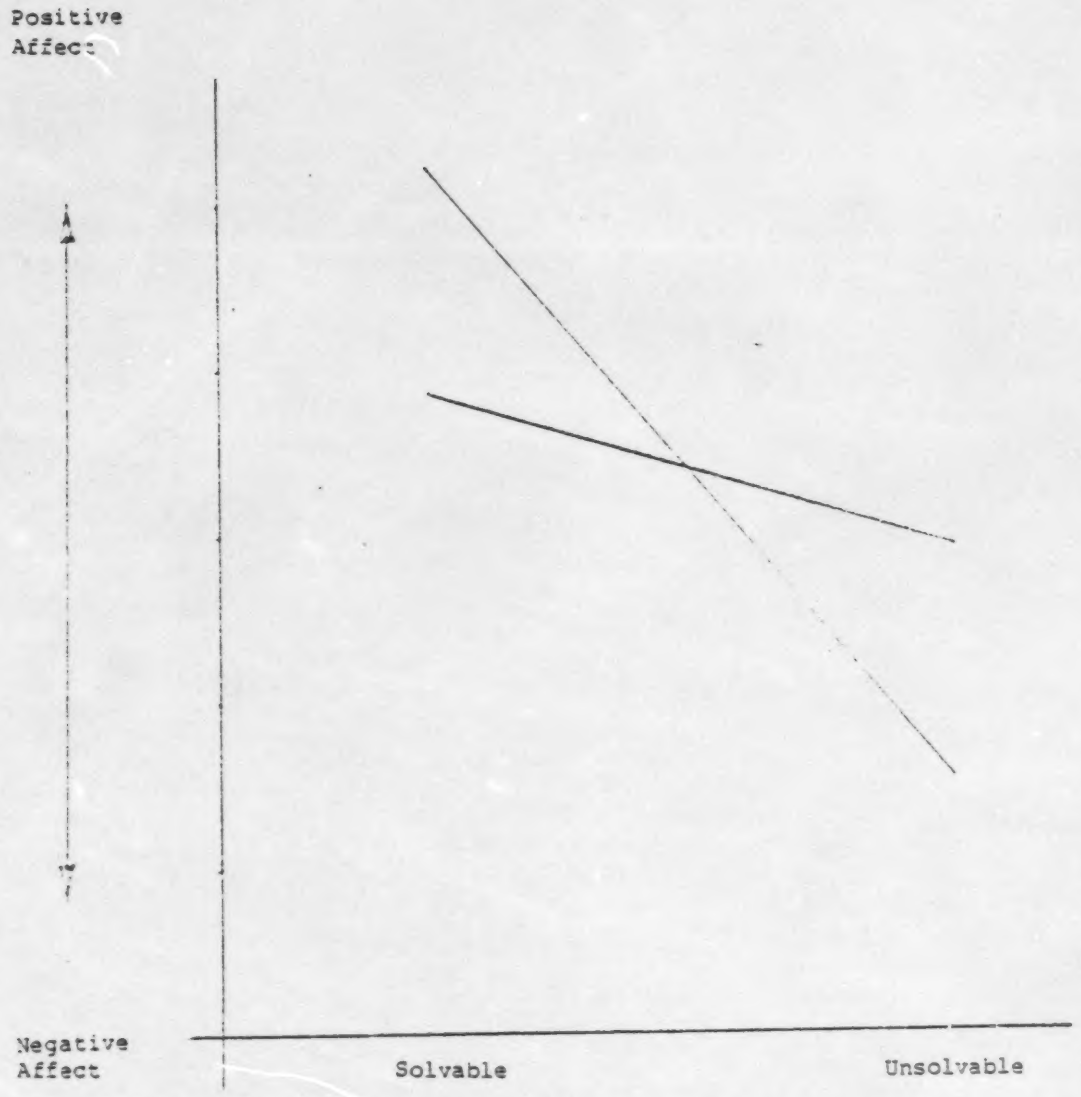
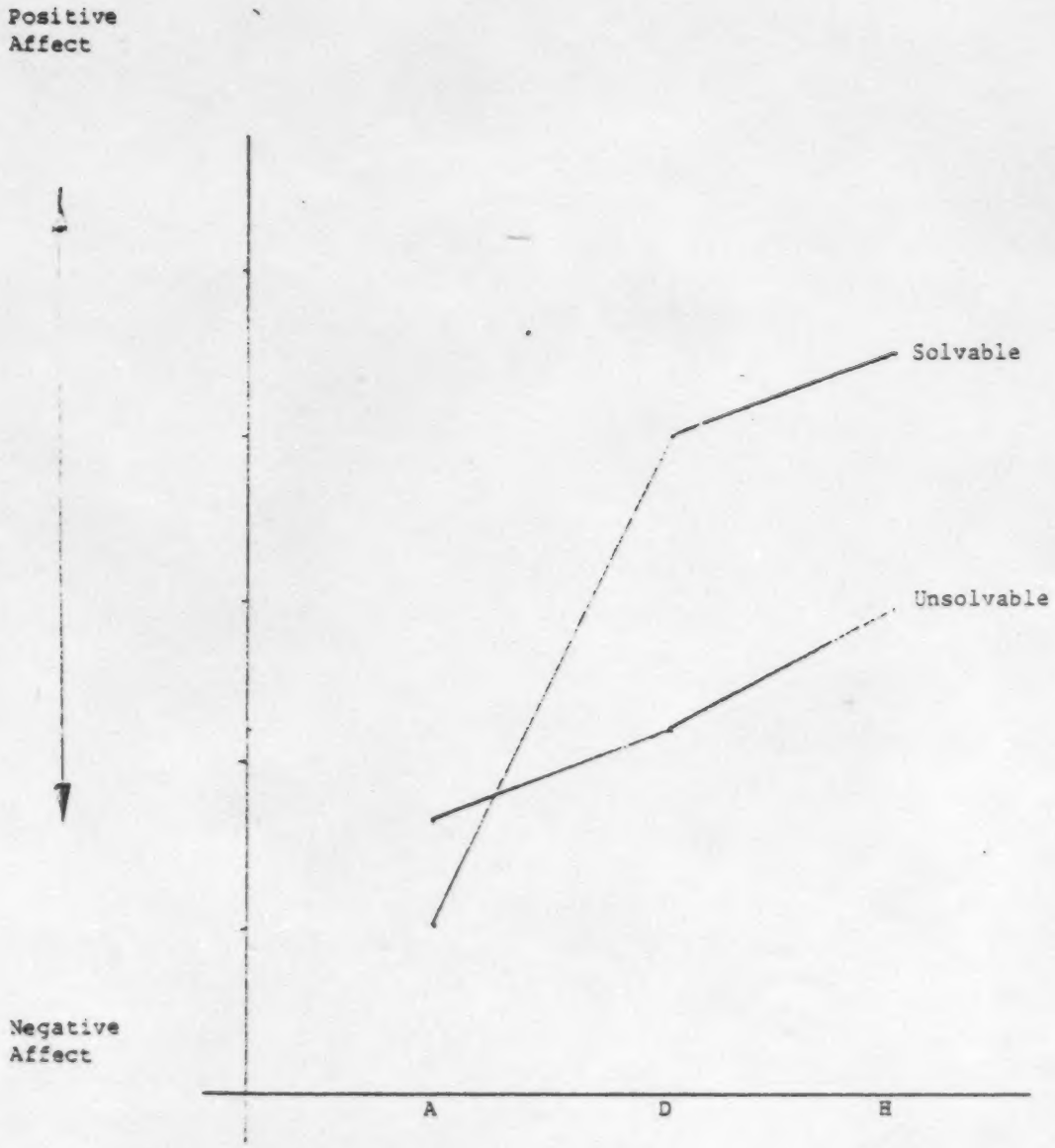


FIGURE 4

Parallelism of Profiles of the Two Solvability Conditions



Summary

The avowed purpose of this paper has been to demonstrate the usage of a variety of post MANOVA data analytic techniques at the disposal of the multivariate researcher. The discerning reader will undoubtedly have noted less than rigorous adherence to some revered principles of experimental design and analysis, such as the strict division of Type I error rates among a family of statistical tests or the tests of validity of distributional assumptions of the data analytic methods. While these considerations are admittedly crucial if strict inferential conclusions are to be made, this paper was written in the spirit of Tukey's (1962) discussion of "The Future of Data Analysis." As paraphrased in the multivariate context by Cooley and Lohnes (1971), the treatment here has been "willing to err moderately often in order that inadequate evidence shall more often suggest the right answer." It has been more interested in "indication procedures" than in "conclusion procedures." Data analysis and interpretation have been the subjects of this paper.

The combination of post MANOVA data analytic strategies can extract much more meaning from the multivariate analysis of data than can simply univariate methods by themselves or even usage of a single multivariate method alone. A brief summary of the contributions of these methods to the analysis of the sample experiment presented here will hopefully demonstrate the advantages of utilizing a number of these post hoc tools to derive the full interpretive benefits of multivariate research.

Discriminant analysis--showed that there was no supplementary discrimination of design effects by the performance and affect subsets of dependent variables when they were analyzed together. It clearly identified the performance variables as detrimental "excess baggage" for one effect (Solvability main effect) and the affect dependent variables as equally useless in another (Attribution main effect), thereby indicating they were more efficiently analyzed as separate outcome packages. Finally, in one case, it verified the identify of the multivariate difference with the one significant univariate result (the Solvability main effect on Depression).

Univariate F tests--identified the single most affected dependent variables for all design effects. In this experiment this was rarely sufficient in and of itself, although it did indicate on a few occasions significant univariate effects which were buried within other nonsignificant effects and masked in the multivariate test. Importantly, the univariate F results always seem to help interpret the meaning of the more multivariate post hoc methods.

Step down F tests--revealed an interesting difference in this example. By removing the differences among Attribution groups on the number of Trials they performed on the experimental task, there were highly significant differences in the number of Failures they experienced. A difference in the number of Failures alone was in no way suggested by the univariate results. This was particularly true of the Internal vs. External contrast.

Profile analysis--indicated that the pattern of Affect for the Solvability groups was quite different. While differences were significant and of similar degrees in ratings of Depression and Hostility, there were virtually no differences at all on Anxiety.

Simultaneous Test Procedures--were used in a modified fashion only. Complex group contrasts were specified a priori and tested on post hoc configurations of dependent variables. Benefits accrued here were in reinforcing the experimental design principle of specifying single degree of freedom contrasts within multiple degrees of freedom effects when they are of interest. Several instances of the tests of Internal vs. External and Stable vs. Unstable attribution contrasts yielded interpretable significance results by themselves where tests of the overall effects (Attribution main effect and S x A interaction) would not have warranted further interpretation.

Univariate multiple comparisons - were illustrated in only one instance in the paper. Hopefully, their place in the post MANOVA data analytic process was sufficiently stated, however, and the reader is directed to the excellent references on these methods provided.

With the array of statistical methods available,, it would seem useful to discuss some general principles for choosing the "best" method of analysis following a significant MANOVA result. While the illustration of these methods may be useful, it is hopefully obvious to the reader that all six presented here speak to their own unique questions of the data. That is, in nearly all instances, if you can specify your post hoc interest clearly, you will have one and only one method suggesting itself. Consistent with the illustrations in this paper and the preceeding summary of the contributions of the six strategies, here are some general descriptions of post hoc interests which would lead to each.

Univariate F tests--when the effects of the treatment on individual dependent variables, without regard to the redundancy or interplay with others are of interest, this is the method of choice.

Discriminant analysis--when the combination of dependent variables which best accounts for treatment differences is of interest, this provides the researcher with information as to how the variables combine in accounting for those differences. Most importantly, it always views each variable in the company of the rest rather than by itself.

Step down F--when any ordering of variables or interest in covarying the effects of one variable from another exist, this method is recommended. According to the ordering of variables presented, it sequentially partitions the multivariate differences into nonoverlapping parts.

Profile analysis--when variables belong to a conceptually common domain and are measured on a comparable scale, this method combines multivariate statistical precision with the visual appeal of examining the patterns of differences across variables.

Simultaneous Test Procedures--allow the researcher to investigate both dimensions of multivariate differences simultaneously--group differences and dependent variable differences. While the other methods deal with one dimension, STP allow both to be investigated in one test.

Univariate multiple comparisons--when the dependent variable differences have been concentrated on one variate (e.g., univariate F, discriminant function composite), these methods are effective in investigating the between groups dimension of the multivariate difference.

References

Anderson, T.W. An introduction to multivariate statistical analysis. New York: Wiley, 1958.

Berger, M.P.L. A note on the use of simultaneous test procedures. Psychological Bulletin, 1978, 85, 895-897.

Bird, K.D. Simultaneous contrast testing procedures for multivariate experiments, Multivariate Behavioral Research, 1975, 10, 343-351.

Bloom, B. S. (Ed.) Taxonomy of educational objectives. Handbook I: cognitive domain. New York: McKay, 1956.

Bock, R.D. Multivariate statistical methods in behavioral research. New York: McGraw-Hill, 1975.

Bock, R.D. and Haggard, E.A. Multivariate analysis of variance. In D.K. Whitla (Ed.) Handbook of Measurement and Assessment in the Behavioral Sciences. Reading, Mass.: Addison-Wesley, 1968.

Cooley, W.W. and Lohnes, P.R. Multivariate data analysis. New York: John Wiley, 1971.

Cramer, E.M. and Bock, R.D. Multivariate analysis. Review of Educational Research, 1966, 36, 604-623.

Darlington, R.B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.

Finn, J.D. MULTIVARIANCE: Univariate and Multivatiate Analysis of Variance Covariance and Regression. A Fortran IV program, Version V, Release 3, 1976.

Finn, J.D. A general model for multivariate analysis. New York: Holt, Rinehart and Winston, 1974.

Gabriel, K.R. A comparison of some methods of simultaneous inference in MANOVA. In P.R. Krishnaiah (Ed.), Multivariate Analysis II. New York: Academic Press, 1969.

Gabriel, K.R. Simultaneous test procedures in multivariate analysis of variance. Biometrika, 1968, 55, 489-504.

Gabriel, R. M. & Glavin, G. B. Multivariate data analysis in empirical research - a look on the bright side. Pavlovian Journal of Biological Science, 1978, 13, 93-112.

Gabriel, R.M. and Hopkins, K.D. Relative merits of MANOVA repeated measures ANOVA; and univariate ANOVAs for research utilizing multiple criterion measures. Journal of Special Education, 1974, 8, 377-389.

Games, P.A. Multiple comparison of means. American Educational Research Journal, 1971, 8, 531-564.

Green, P.E. Analyzing multivariate data. Hinsdale, Ill.: Dryden Press, 1978.

Harris, R.J. A primer of multivariate statistics. New York: Academic Press, 1975.

Heck, D.L. Charts of some upper percentage points of the distribution of the largest characteristic root. Annals of Mathematical Statistics, 1960, 31, 625-642.

Hopkins, K.D. & CHadbourn, R.A. A scheme for proper utilization of multiple comparisons in research and a case study. American Educational Research Journal, 1967, 4, 407-412.

Huberty, C.J. Discriminant Analysis. Review of Educational Research, 1975, 45, 543-598.

Huberty, C.J. The stability of three indices of relative variable contribution in discriminant analysis. Journal of Experimental Education, 1975, 44, 59-64.

Hummel, T.J. & Sligo, J.R. Empirical comparison of univariate and multivariate analysis of variance procedures. Psychological Bulletin, 1971, 76, 49-57.

Kaplan, R.M. & Litrownik, A.J. Some statistical methods for the assessment of multiple outcome criteria in behavioral research. Behavior Therapy, 1977, 8, 383-392.

Kirk, R.E., Experimental Design: procedures for the behavioral sciences, Monterey, California: Brooks/Cole, 1968.

Miller, R.G. Simultaneous statistical inference. New York: McGraw-Hill, 1966.

Morrison, D.F. Multivariate statistical methods, New York: McGraw-Hill, 1976.

Olson, C.L. Comparative robustness of six tests in multivariate analysis of variance. Journal of the Americal Statistical Association, 1974, 69, 894-908.

Olson, C.L. On choosing a test statistic in multivariate analysis of variance. Psychological Bulletin, 1976, 83, 579-586.

Overall, J.E. and Klett, C.J. Applied multivariate analysis. New York: McGraw-Hill, 1972.

Press, S.J. Applied multivariate analysis. New York: Holt Rinehart and Winston, 1972.

Rao, C.R. Advanced statistical methods in biometric research. New York: Wiley & Sons, 1952.

Rao, C.R. Linear statistical inference and its applications. New York: Wiley, 1965.

Spector, P.E. What to do with significant multivariate effects in multivariate analysis of variance. Journal of Applied Psychology, 1977, 62, 158-163.

Stevens, J.P. Four methods of analyzing between variation for the k-group MANOVA problem. Multivariate Behavioral Research, 1972, 7, 499-501.

Stevens, J.P. Step down analysis and simultaneous confidence intervals in MANOVA. Multivariate Behavioral Research, 1973, 8, 391-402.

Tatsuoka, M.M. Discriminant Analysis: the study of group differences. Champaign, Illinois: Institute for Personality and Ability Testing, 1970.

Tatsuoka, M.M. Significance Tests: univariate and multivariate. Champaign, Illinois: Institutue for Personality and Ability Testing, 1971 (a).

Tatsuoka, M.M. Multivariate analysis: techniques for educational and psychological research. New York: John Wiley and Sons, 1971 (b).

Tatsuoka, M.M. Multivariate analysis in educational research. In F.N. Kerlinger (Ed.). Review of Research in Education. Itasca, Illinois: F. E. Peacock, 1973.

Timm, N.H. Multivariate analysis: with applications in education and psychology. Monterey, California: Brooks/Cole, 1975.

Tukey, J.W. Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley, 1977.

Tukey, J.W. The future of data analysis. Annals of Mathematical Statistics, 1962, 33, 1-67.

Wilkinson, L. Response variable hypotheses in the multivariate analysis of variance. Psychological Bulletin, 1975, 82, 408-412.

Wilks, S.S. Certain generalizations in the analysis of variance. Biometrika, 1932, 24, 471-494.